心理学报 2018, Vol. 50, No.12, 1346−1355
Acta Psychologica Sinica DOI: 10.3724/SP.J.1041.2018.01346

歌词对音乐情绪加工的影响：行为与 ERP 研究*

张伟霞[1] 王莞琪[2] 周临舒[3] 蒋存梅[3]

([1]上海师范大学教育学院; [2]上海师范大学商学院; [3]上海师范大学音乐学院, 上海 200234)

摘 要 本研究探讨了歌词对音乐情绪加工的影响。实验 1 使用情感启动范式, 带有歌词与无歌词音乐片段为启动刺激, 与音乐情绪一致或不一致的面孔图片为目标刺激, 被试任务是既快又准确地判断目标面孔的情绪。结果显示, 无论音乐是否带有歌词, 听者在一致条件下的反应都比不一致条件更快更准确, 这表明听者能加工音乐传达的情绪信息。实验 2 进一步通过电生理手段探讨歌词影响音乐情绪加工的神经机制。研究结果显示, 尽管听者对带有歌词和无歌词音乐情绪的加工都产生了启动效应, 但是无歌词音乐条件在 250~450 ms 时间窗口产生了 N400 效应, 而带有歌词音乐条件在 500~700 ms 时间窗口诱发了 LPC 效应, 该结果表明, 歌词影响了大脑加工音乐情绪的时间进程。本研究结果将在一定程度上为音乐与语言关系的探究提供依据。

关键词 音乐情绪; 语言; 歌词; N400; LPC

分类号 B842

1 前言

音乐和语言是人类社会独有的，属于人类交流的两种重要手段。早在 1871 年, Darwin 就提出了原始母语(protolanguage)假说，认为语言和音乐可能有相同的起源。对于远古人类来说，语言和音乐的主要功能可能在于情绪表达(Thompson, Marin, & Stewart, 2012)。随着人类社会化程度的提高，语言和音乐逐渐分化，并朝着不同的进化方向发展(Mithen, 2006; Perlovsky, 2010)。语言发展成为具有明确语义的符号交流系统，音乐则成为情绪表达的重要手段(Jackendoff, 2009)。

音乐可能先于语言产生，在语言形成之前，人类可能就能够通过类似音乐的方式进行交流(Darwin, 1871; Levman, 1992)。那么，具有明确语义的语言信息是否有助于音乐的情绪表现？或者说，音乐作为一种独立的交流手段，其情绪表现并不依赖于语义信息，语义信息反而影响其情绪表现？歌曲包含歌词和旋律，兼具语言和音乐的特征。在音乐作品中，演唱者可以仅呈现歌曲的旋律(比如，以无意义音节“la”演唱)，也可以同时呈现旋律与歌词。因此，已有很多研究从歌曲入手，试图揭示具有明确语义的语言信息对音乐情绪加工的影响，该研究问题将有助于阐明语言与音乐在情绪加工方面的关系。

已有研究主要从行为层面对此进行了探讨。研究表明，歌词的确影响了听者对音乐情绪的加工(Ali & Peynircioğlu, 2006; Mori & Iwanaga, 2013; Stratton & Zalanowski, 1994)。比如，Stratton 和 Zalanowski (1994)以歌曲《Why I was born》为原始刺激，制作了纯歌词、纯音乐以及歌词加钢琴伴奏三种版本。实验中要求听者评价刺激的愉悦度，并通过量表测量听者在实验前后的情绪变化。结果表明，听者对无歌词版本的音乐愉悦度评价显著高于

收稿日期: 2018-03-02
* 国家自然科学基金项目(31470972, 31500876)资助。
张伟霞和王莞琪为共同第一作者。
通信作者: 蒋存梅, E-mail: cunmeijiang@126.com

带有歌词的两个版本。无歌词音乐使得听者的负性情绪减少，正性情绪增强。带有歌词的音乐则呈现出完全相反的趋势：听者的负性情绪增强而正性情绪减弱。该结果表明，无论歌词是独立呈现，还是伴随旋律呈现都能影响听者对音乐的情绪加工。歌词对音乐情绪加工的影响也得到了计算机算法研究的证实。研究发现，算法中是否包含歌词信息会影响该算法对音乐情绪自动分类的准确率(Laurier, Grivolla, & Herrera, 2008)及击中率(Hu, Downie, & Ehmann, 2009)。

然而，也有研究发现，无论歌词的表现形式是说出来，还是唱出来，听者在聆听带有歌词音乐前后的情绪体验并没有发生非常显著的改变(Galizio & Hendrick, 1972)。类似地，通过测量听者聆听音乐前后的心境变化，Sousou (1997)发现，歌词并不影响听者对音乐的情绪体验。已有研究结果之间的差异可能来源于两个方面。一方面是实验中对比条件的设置存在差异。具体来说，Galizio 和 Hendrick (1972)与 Sousou (1997)的研究并未设置无歌词的对比条件，而另一些研究(Ali & Peynircioğlu, 2006; Brattico et al., 2011; Stratton & Zalanowski, 1994)均设置了无歌词条件。另一方面可能在于，研究者没有使用客观的指标量化听者对音乐情绪的加工。在以上研究中，研究者大都使用了主观评定的方法。比如，Stratton 和 Zalanowski (1994)使用多元情感形容词检核表修订版(Multiple Affect Adjective Check List-R)，而 Sousou (1997)与 Mori 和 Iwanaga (2013)则使用自编的等级评定条目；同时，各研究的评定内容从愉悦度(Stratton & Zalanowski, 1994; Mori & Iwanaga, 2013)到强度(Ali & Peynircioğlu, 2006)也不尽相同。这些都导致了已有研究结果之间对比的困难。

基于此，本研究试图在行为研究的基础上，通过电生理手段探究歌词对音乐情绪加工的影响。实验 1 通过认知行为方法考察听者能否加工有/无歌词音乐所传达的情绪。如果实验 1 发现，无论音乐是否带有歌词，听者都能加工其情绪，实验 2 则进一步通过电生理手段考察大脑加工这两类音乐所传递的情绪是否存在差异。两个实验均采用情感启动范式，音乐刺激都是 120 条由声乐家演唱的音乐片段，形成两种条件：带有歌词条件，即演唱中文歌词；无歌词条件，即以无意义音节“la”演唱。这两种条件的音乐刺激都作为启动刺激，与音乐情绪效价一致或不一致的面孔图片作为目标刺激。如果音乐能够启动听者对情绪面孔的加工，则意味着听者能加工音乐所传达的情绪信息。如果带有歌词条件下的音乐情绪加工模式与无歌词条件下存在差异，则意味着歌词对音乐情绪加工具有影响。

2　实验 1：歌词影响音乐情绪加工的行为研究

实验 1 从行为层面上探讨听者能否加工带有歌词与无歌词音乐的情绪信息。为了排除熟悉性的影响，本实验的音乐刺激均选自欧洲歌剧片段。这是因为，本研究所招募的被试均是未受过专业音乐训练的普通大学生。对中国大学生而言，他们日常主要聆听流行音乐，较少主动接触欧洲歌剧。此外，为排除音色的影响，本实验的音乐刺激均由人声演唱。在无歌词版本中，演唱者以无意义音节“la”演唱；在带有歌词版本中，演唱者以中文歌词(公开出版的中译本)演唱。为了确保两个演唱版本的有效性，我们实施了 3 个前测(具体内容见下文刺激与程序部分)。如果带有歌词和无歌词音乐都能启动听者对情绪面孔的加工，那么，说明听者能够加工这两种音乐所传达的情绪信息。

2.1　方法

2.1.1　被试

40 名未接受专业音乐训练的普通大学生自愿参加了本实验。所有被试均为右利手，听力正常，视力或矫正视力正常，无神经或精神方面的病史，且在实验前签署了知情同意书。因为错误理解指导语，8 名被试的数据被剔除。最终剩余 32 名有效被试(24.47 ± 1.65 岁，15 男)。

2.1.2　刺激与程序

120 条原始音乐刺激均选自欧洲歌剧片段，表达高兴与悲伤情绪的音乐各半。所有音乐刺激均由一名声乐表演者(接受过 18 年专业的美声歌唱训练)演唱。每一原始音乐片断均用中文歌词及无意义音节“la”演唱，分别形成带有歌词和无歌词音乐两个版本。音乐录制前，要求该声乐表演者仔细斟酌每一音乐片段的表演方式，力求两个版本在力度(即表演者对音响强弱变化的处理方式)、弹性速度(rubato，即表演者在一定范围内对速度自由变化的把握)、分句(即乐句的断句或演唱的呼吸口)及整体表演水平(即音响整体质量的好坏)等方面的一致性。录制后的音乐统一使用 Adobe Audition CS6 (Adobe Systems Inc)和 Goldwave 剪辑处理。音乐刺激平均时长 17 s (范围：10~25 s)，皆为单频道，采

样率 22.050 kHz, 16 位分辨率, 平均响度均标准化为–7 dB, 同时淡出 1 s。

为确保实验材料的有效性, 本实验实施了三个前测。第一个前测是为了保证两个版本的音乐刺激在表演水平和表演方式上不存在显著差异。我们招募 16 名音乐专业大学生(均接受了 18 年专业音乐训练), 要求其评定带有歌词与无歌词音乐在弹性速度, 力度, 分句处理方式以及总体表演水平上的一致性(1 = 非常不同, 4 = 不确定, 7 = 非常相同)。只有平均分高于 4 的音乐刺激才被采用。第二个前测是为了确保听者能听清带有歌词音乐中的歌词内容。我们招募 16 名不参与正式实验的普通大学生, 要求其评定音乐中歌词的清晰度(1 = 不清楚, 3 = 不确定, 5 = 清楚)。只有平均分高于 4 的音乐刺激才被采用。经过两个前测筛选, 带有歌词与无歌词音乐刺激中各有 80 条符合要求。第三个前测是为了确保启动音乐和目标面孔在情绪效价上的关系是一致或不一致。我们从《中国化面孔情绪图片系统》(龚栩, 黄宇霞, 王妍, 罗跃嘉, 2011)中选取悲伤和高兴面孔各 80 张, 作为潜在的目标刺激, 男女面孔各半。通过 Adobe Photoshop CS 调整面孔图片像素为 102 × 768, 16 位分辨率。160 条音乐刺激均呈现两次, 分别和情绪一致及情绪不一致的面孔匹配, 形成 320 对音乐–面孔配对。招募 16 名不参加正式实验的普通大学生, 要求其评定音乐–面孔配对的一致性(1 = 非常不一致, 5 = 不确定, 9 = 非常一致)。得分高于 7 的配对视作情绪一致的刺激, 得分低于 3 的配对视作情绪不一致的刺激。最终带有歌词音乐和无歌词音乐启动条件下各有 60 条刺激符合标准, 形成 4 种条件: 带有歌词音乐–图片一致、带有歌词音乐–图片不一致、无歌词音乐–图片一致, 以及无歌词音乐–图片不一致(详见图 1), 每个实验条件各有 60 个配对。我们进一步对以上配对的情绪一致性评定结果进行 2(一致性: 一致, 不一致) × 2(歌词: 带有歌词音乐, 无歌词音乐)重复测量方差分析。结果表明, 一致性主效应显著, $F(1, 59) = 2318.45$, $p < 0.001$, $\eta_p^2 = 0.98$。无论带有歌词音乐(一致: M = 7.37, SD = 0.53; 不一致: M = 2.70, SD = 0.57), 还是无歌词音乐(一致: M = 7.33, SD = 0.55; 不一致: M = 2.66, SD = 0.56), 听者对一致的评定均显著高于不一致。其它效应均不显著($ps > 0.09$)。该研究结果表明, 无论音乐刺激是否带有歌词, 情绪一致与不一致条件均存在差异。此外, 我们还计算了第 3 个前测中筛选出的带有歌词音乐的清晰度: 歌词的清晰度平均得分为 4.42 (SD = 0.23); 带有歌词音乐与无歌词音乐在表情因素方面的处理方式以及总体表演水平的一致性得分如下: 弹性速度(5.44 ± 0.48)、力度(5.56 ± 0.47)、分句(5.56 ± 0.47)、总体表演水平(5.52 ± 0.44)。

我们将 240 个配对以拉丁方的方式分配到 2 套刺激中, 每套刺激各有 120 个试次, 包含 4 个实验条件, 每个被试只需完成其中一套刺激。刺激以伪随机方式呈现, 同一启动或目标刺激之间至少间隔 8 个试次。在实验中, 1000 ms 黑色注视点消失后通过飞利浦头戴式 SHM1900 耳机播放启动音乐。音乐播放完毕后呈现情绪面孔图片。被试任务是

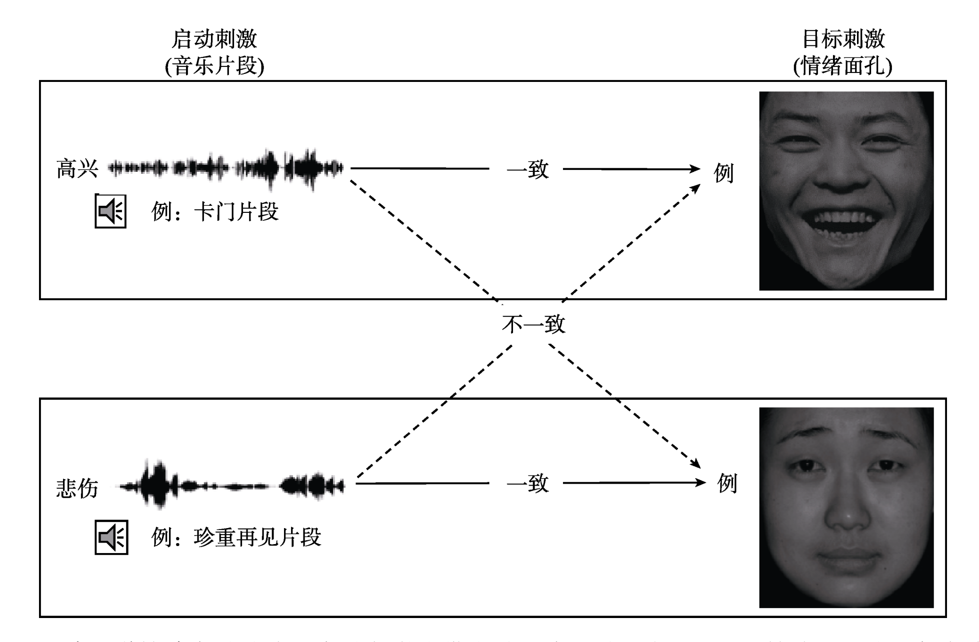


图 1 跨通道情感启动范式。启动刺激为带有歌词或无歌词音乐片段, 情绪面孔为目标刺激

又快又准地判断图片情绪, 高兴按 F 悲伤按 J。情绪类型(高兴/悲伤)与按键(F/J)的对应关系在被试间平衡。判断结束后, 按空格键开始下一个试次。正式实验开始前, 我们提供 4 个练习试次以确保被试熟悉实验流程。为进一步排除熟悉的影响, 正式实验结束后, 我们要求被试报告实验中音乐片段对应的标题。如果被试能报告出标题中一个以上的关键词, 表明被试熟悉该音乐片段。在该实验中, 没有被试能报告出所有音乐名称中的任何一个关键词。

2.2　结果与讨论

图 2 显示了 4 种实验条件下的平均正确率(a)和反应时(b)。在正确率方面, 2(歌词: 带有歌词音乐, 无歌词音乐) × 2(一致性: 一致, 不一致)重复测量方差分析结果发现: 一致性主效应显著, $F(1, 31) = 9.80$, $p = 0.004$, $\eta_p^2 = 0.24$, 表明一致条件(M = 97.86%, SD = 2.74%)的正确率显著高于不一致条件(M = 94.71%, SD = 6.79%); 歌词主效应显著($F(1, 31) = 7.64$, $p = 0.01$, $\eta_p^2 = 0.20$), 表明无歌词音乐条件(M = 97.17%, SD = 4.03%)的正确率显著高于带有歌词音乐条件(M = 95.40%, SD = 6.39%), 交互作用不显著($p > 0.05$)。

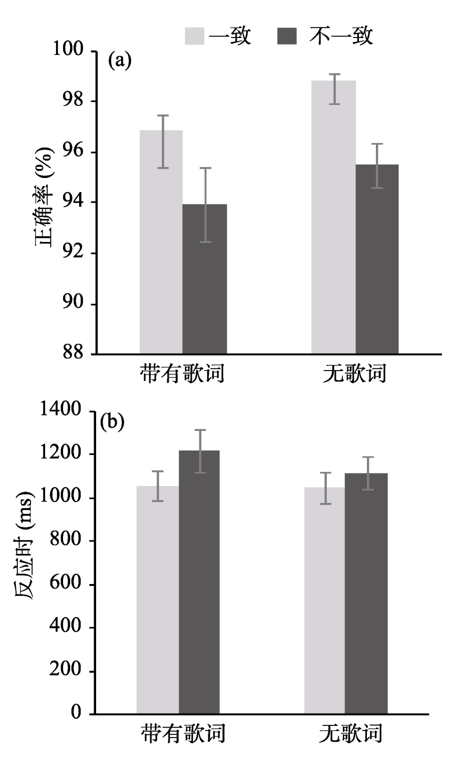


图 2　每种实验条件下的平均正确率(a)和反应时(b), 误差线为标准误

在反应时方面, 2(歌词: 带有歌词音乐, 无歌词音乐) × 2(一致性: 一致, 不一致)重复测量方差分析结果表明: 一致性主效应显著($F(1, 31) = 14.38$, $p = 0.001$, $\eta_p^2 = 0.32$), 情绪一致条件下的反应时(M = 1051.83 ms, SD = 389.35 ms)显著低不一致条件(M = 1166.45 ms, SD = 503.22 ms); 歌词主效应显著($F(1, 31) = 4.42$, $p = 0.04$, $\eta_p^2 = 0.13$), 带有歌词音乐条件下的反应时(M = 1136.59 ms, SD = 486.83 ms)显著长于无歌词音乐条件(M = 1081.69 ms, SD = 415.88)。歌词与一致性交互作用不显著($p > 0.05$)。

与已有研究的行为结果一致(Goerlich, Witteman, Aleman, & Martens, 2011; Wang & Qin, 2016; Zhang, Li, Gold, & Jiang, 2010), 在本研究中, 相对于情绪不一致条件, 听者在情绪一致条件下的反应更快更准。这说明, 带有歌词和无歌词音乐都能启动听者对情绪面孔的加工。也就是说, 听者能加工带有歌词和无歌词音乐的情绪(Morton & Trehub, 2007)。

3　实验 2: 歌词影响音乐情绪加工的 ERP 研究

实验 1 结果表明, 无论是否带有歌词, 听者都能够加工音乐的情绪信息。由于行为实验无法反映出大脑加工音乐情绪在时间进程上的动态变化, 因此, 实验 2 试图通过电生理手段进一步探究歌词影响音乐情绪加工的神经机制。在已有研究中, 仅有一个研究借助功能核磁共振成像(fMRI)技术探讨了歌词影响音乐情绪加工的神经机制, 遗憾的是, 研究者使用带有歌词的声乐曲和无歌词的器乐曲为实验刺激, 这种实验操纵无法排除音色在音乐情绪加工中的作用(Brattico et al, 2011)。的确, 众多研究表明, 音色对音乐情绪加工具有重要的影响(Behrens & Green, 1993; Hailstone et al., 2009; Franco, Chew, & Swaine, 2017)。那么, 控制音色之后, 歌词是否影响听者对音乐情绪的神经加工? 这是本实验要探究的问题。

先前使用情感启动范式的研究主要关注 N400 与晚期正成分(late positive component, LPC)。研究表明, 情绪不一致条件比一致条件诱发了更大的 N400, 该成分反映了大脑对情绪冲突的检测(Schirmer, Kotz, & Friederici, 2002; Zhang, Lawson, Guo, & Jiang, 2006)以及对情绪信息的整合加工(Kamiyama, Abla, Iwanaga, & Okanoya, 2013; Zhang et al., 2010)。类似地, 情绪不一致条件诱发的 LPC 比情绪一致条件更大(Herring, Taylor, White, & Crites, 2011; Werheid, Alpay, Jentzsch, & Sommer,

2005; Zhang et al., 2010)，表明大脑对情绪不一致条件的加工需要更多的注意参与(Zhang, Kong, & Jiang, 2012; Zhang et al, 2010)。如果听者对带有歌词音乐和无歌词音乐所传达的情绪信息都能进行加工(Morton & Trehub, 2007)，且歌词会影响听者对音乐情绪的加工(Ali & Peynircioğlu, 2006; Stratton & Zalanowski, 1994)。我们预期，听者对带有歌词音乐和无歌词音乐情绪信息的加工都会产生启动效应，也就是 N400 或 LPC 效应。但是，有歌词音乐和无歌词音乐启动条件下所产生的启动效应可能存在差异。

3.1 方法

3.1.1 被试

被试是 20 名没有受过专业音乐训练的普通大学生，4 名被试因脑电伪迹过多被剔除，有效被试为 16 名(23.88 ± 1.36 岁，7 男)。所有被试均为右利手，听力正常，视力或矫正视力正常，无精神病史，无大脑损伤。所有被试均签署了知情同意书，并在实验结束后获得一定的报酬。

3.1.2 刺激与程序

刺激材料与实验 1 相同。为避免行为反应对脑电信号的干扰，实验 2 在程序上与实验 1 略有不同。在实验 2 中，音乐播放完毕后呈现的情绪面孔图片不是按键消失，而是固定的 1000 ms。事件相关脑电的标记打在情绪面孔呈现的时间点上。在面孔消失后立即出现反应界面，要求被试判断音乐与面孔所表达的情绪是否一致，一致按 F 不一致按 J。一致性(一致/不一致)与按键(F/J)的对应关系在被试间平衡。反应时间没有限定，判断结束后，按空格键开始下一个试次。正式实验开始前，被试完成 6 个练习刺激以熟悉实验流程。与实验 1 相同，为进一步排除熟悉度可能对实验造成的干扰，在正式实验结束后，要求被试报告实验中音乐片段的标题。在该实验中，没有被试能报告出所有音乐标题中的任何一个关键词。

3.1.3 脑电记录与数据分析

我们采用 Biosemi 64 导 Active Two 电极帽，在 2048 Hz 采样率下记录 EEG 信号。外接电极分别放置于左、右外眼角处及左眼上、下以分别记录水平眼电和垂直眼电。脑电采集中控制电极与头皮接触的电阻在 20 kΩ 以下。离线分析时，将参考转为双侧乳突平均参考，进行 0.1~30 Hz (24 dB/oct)带通滤波，并通过 BESA 分析软件自动矫正眼电伪迹。分段为目标刺激呈现前 200 ms 到目标刺激呈现后 1000 ms，基线为目标刺激出现前的 200 ms 时间段。删除波幅变化超过±120 μV 及反应错误的试次。

根据半球和区域，我们选取 9 个感兴趣区(interest of region, ROI)的电极点(左前：FP1, AF3, F3, F5, F7；中前：FPz, AFz, Fz；右前：FP2, AF4, F4, F6, F8；左中：C1, CP1, FC3, C3, CP3；中中：FCz, Cz, CPz；右中：C2, CP2, FC4, C4, CP4；左后：P3, P5, PO3, PO7, O1；中后：Pz, POz, Oz；右后：P4, P6, PO4, PO8, O2)分别进行中线和两侧的重复测量方差分析。就中线电极点而言，一致性(一致，不一致)，歌词(带有歌词音乐，无歌词音乐)和脑区(前，中，后)为被试内因素。两侧分析在中线分析的基础上增加了半球(左，右)作为被试内因素。计算每个兴趣区内所有电极点的平均值并做进一步分析。本研究只报告显著或边缘显著的主要实验变量(歌词，一致性)的统计结果。交互作用显著后进行简单效应分析，所有的成对比较均使用 Bonferroni correction 矫正。球形假设不成立时，使用 Greenhouse-Geisser 矫正 *p* 值。

3.2 结果与讨论

3.2.1 行为结果

我们以正确率为因变量，进行 2(歌词：带有歌词音乐，无歌词音乐) × 2(一致性:一致，不一致)的重复测量方差分析。结果没有发现任何显著的效应($ps > 0.05$)。带有歌词条件下的正确率为 85.83%，无歌词条件下的正确率为 83.02%，这表明，无论是带有歌词条件还是无歌词条件，听者都能认真完成实验任务。

3.2.2 脑电结果

图 3 呈现了无歌词(a)与带有歌词(b)音乐启动条件下产生的 ERP 波形图。图 4 呈现的是在 250~450 ms (a)和 500~700 ms (b)时间窗口内无歌词与带有歌词音乐条件下的差异波(情绪不一致减情绪一致)地形图。基于图形观察，以及已有研究发现(Daltrozzo & Schön, 2009; Herring et al., 2011; Kamiyama et al., 2013; Werheid et al., 2005; Zhang et al., 2010)的研究，我们将目标刺激出现后的 250~450 ms 及 500~700 ms 分别作为 N400 和 LPC 的时间窗。

我们以 N400 平均波幅为因变量，分别进行中线和两侧分析。中线分析中，2(歌词：带有歌词音乐，无歌词音乐) × 2(一致性：一致，不一致) × 3(脑区：前，中，后)三因素重复测量方差分析发现，歌词与一致性交互作用显著，$F(1, 15) = 8.48$, $p = 0.01$, $\eta_p^2 =$

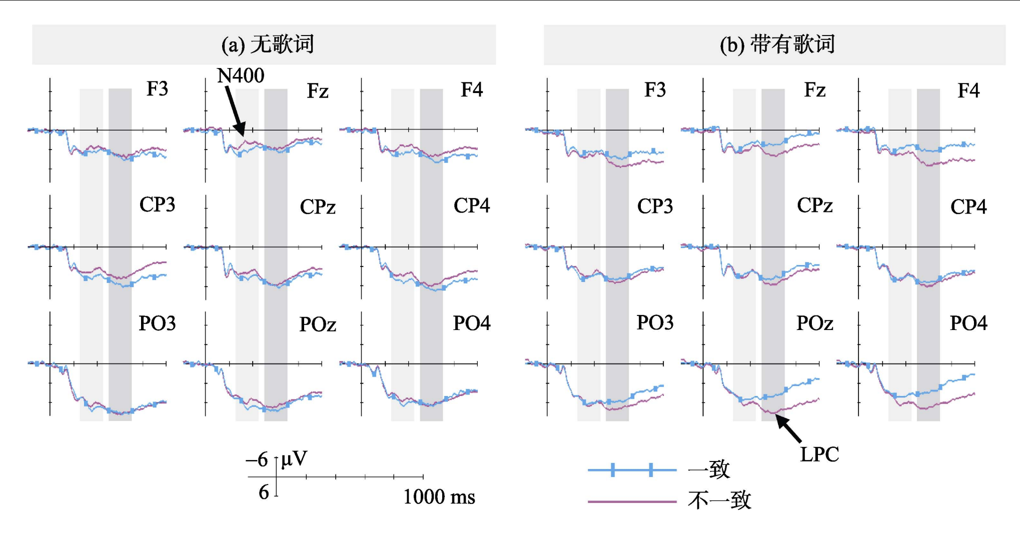


图 3　无歌词(a)与带有歌词(b)启动条件下的 ERP 波形图。浅灰与中灰阴影部分分别表示 N400 与 LPC 的时间窗口

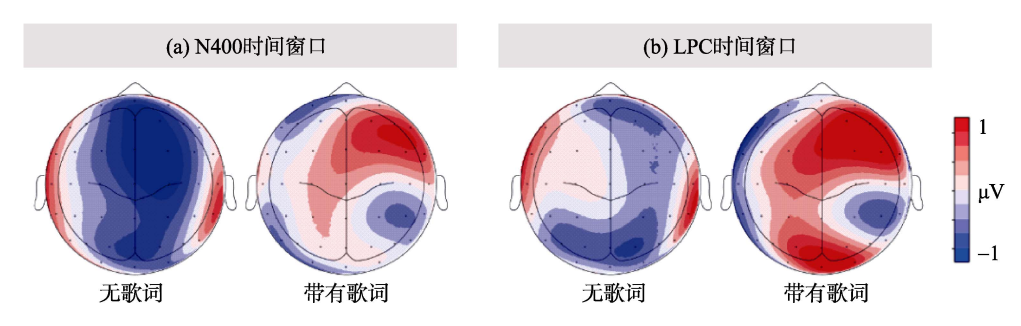


图 4　带有歌词与无歌词在 250~450 ms (a)及 500~700 ms (b)时间窗内的差异波地形图
注：彩图见电子版

0.36。简单效应分析表明，在无歌词音乐启动下，情绪不一致条件比一致条件诱发了更大的 N400 波幅，$F(1, 15) = 5.17$, $p = 0.04$, $\eta_p^2 = 0.26$。然而，在带有歌词音乐启动下，情绪一致与不一致条件诱发的 N400 波幅不存在显著差异，$F(1, 15) = 2.90$, $p = 0.11$。在两侧分析中，2(歌词：带有歌词音乐，无歌词音乐) × 2(一致性：一致，不一致) × 3(脑区：前，中，后) × 2(半球：左，右)四因素重复测量方差分析发现，歌词与一致性的交互作用显著，$F(1, 15) = 7.80$, $p = 0.02$, $\eta_p^2 = 0.34$。简单效应分析表明，在无歌词音乐启动下，情绪不一致条件比一致条件诱发了更大的 N400 波幅，$F(1, 15) = 6.81$, $p = 0.02$, $\eta_p^2 = 0.31$。然而，在带有歌词音乐启动下，情绪一致与不一致条件诱发的 N400 波幅不存在显著差异，$F(1, 15) = 2.18$, $p = 0.16$。在中线和两侧分析中，其它与歌词或一致性有关的效应均不显著($ps > 0.35$)。

类似地，我们以 LPC 平均波幅为因变量，分别进行中线和两侧的分析。在中线分析中，2(歌词：带有歌词音乐，无歌词音乐) × 2(一致性：一致，不一致) × 3(脑区：前，中，后)三因素重复测量方差分析表明，歌词与一致性交互作用显著，$F(1, 15) = 7.47$, $p = 0.02$, $\eta_p^2 = 0.33$。简单效应分析表明，在带有歌词音乐启动下，情绪不一致条件比一致条件诱发了更大的 LPC 波幅，$F(1, 15) = 6.90$, $p = 0.02$, $\eta_p^2 = 0.32$。在无歌词音乐启动下，情绪不一致条件与一致条件诱发的 LPC 波幅不存在显著差异，$F(1, 15) = 0.78$, $p = 0.39$。在两侧分析中，2(歌词：带有歌词音乐，无歌词音乐) × 2(一致性：一致，不一致) × 3(脑区：前，中，后) × 2(半球：左，右)四因素重复测量方差分析发现，歌词与一致性交互作用显著，$F(1, 15) = 6.20$, $p = 0.03$, $\eta_p^2 = 0.29$。简单效应分析表明，在带有歌词音乐启动下，情绪不一致条件比一致条件诱发了更大的 LPC，$F(1, 15) = 5.17$, $p = 0.04$, $\eta_p^2 = 0.26$。在无歌词音乐启动下，情绪不一致条件与一致条件诱发的 LPC 波幅不存在显著差异，$F(1, 15) = 0.82$, $p = 0.38$。中线和两侧分析中没有其它与歌词或一致性有关的显著效应($ps > 0.21$)。

实验 2 脑电结果发现，带有歌词音乐启动条件下在 250~450 ms 时间窗口诱发了 N400 效应，无歌词音乐启动条件下在 500~700 ms 时间窗口诱发了

LPC 效应。该结果表明，无论音乐是否带有歌词，听者均能对启动刺激与目标刺激之间的情绪关系进行判断，但是，听者加工带有歌词与无歌词音乐情绪存在时间进程上的差异。也就是说，听者对带有歌词音乐情绪的加工滞后于无歌词音乐。

4 综合讨论

本研究通过两个实验探讨歌词影响音乐情绪加工的神经机制。行为结果显示，无论音乐是否带有歌词，听者在一致条件下的反应都比不一致条件更快更准确，表明无论音乐是否带有歌词，听者都能加工音乐传达的情绪信息。ERP 结果进一步显示，尽管听者对带有歌词与无歌词音乐情绪的加工都诱发了启动效应，但是无歌词音乐条件在 250~ 450 ms 时间窗口诱发了 N400 效应，而带有歌词音乐条件在 500~700 ms 时间窗口诱发了 LPC 效应，该结果表明，歌词影响了听者加工音乐情绪的时间进程，说明听者对带有歌词音乐情绪的加工滞后于无歌词音乐。

经典 N400 通常和语义加工有关，不一致的语义常常比一致诱发更大的 N400 (Kutas & Hillyard, 1980)，该成分体现出大脑对概念的整合加工(Brown & Hagoort, 1993; Kutas & Federmeier, 2000; Kutas & Federmeier, 2011)。近年研究也表明，在情感启动范式中，情绪不一致的条件比情绪一致的条件诱发了更大的 N400 (Schirmer et al., 2002; Zhang et al., 2010; Zhang et al., 2006)，表示对情绪信息的整合加工需要更多的认知资源(Kamiyama et al, 2013; Zhang et al, 2010)。研究也发现，当启动刺激为短小音乐片段(Daltrozzo & Schön, 2009; Goerlich et al., 2011; Koelsch et al., 2004)与和弦(Steinbeis & Koelsch, 2011)时，大脑也能对不一致的情绪信息诱发更大波幅的 N400。本研究发现，对无歌词音乐情绪的加工诱发了 N400 效应，这一效应的产生主要源于情绪意义的激活(Daltrozzo & Schön, 2009; Eder, Leuthold, Rothermund, & Schweinberger, 2011)。的确，启动刺激的出现在概念水平上预先激活了与目标刺激有关的情绪表征，减少了情绪一致条件下的 N400 波幅(Goerlich et al., 2012)。相比情绪一致条件，在情绪不一致条件下，由于与目标刺激相关的情绪表征没有被提前激活，所以听者对无歌词音乐情绪信息与面孔情绪信息的整合加工需要更多的认知资源，体现为 N400 波幅的增大。

与以往使用情感启动范式的研究一致(Herring et al., 2011; Wang & Qin, 2016; Werheid et al., 2005; Zhang et al., 2012; Zhang et al., 2010)，在本研究中，听者对带有歌词音乐的情绪加工诱发了 LPC 效应。这可能反映了注意资源的分配。的确，有研究发现情感启动范式中不一致条件下更大的 LPC 是由于更多的注意卷入所致(Herring et al., 2011; Werheid et al., 2005; Zhang et al., 2010)。另一方面，已有研究发现 LPC 体现了整合加工(Juottonen, Revonsuo, & Lang, 1996)。在语言与音乐的记忆研究中，研究者发现，当语言与旋律信息同时呈现时，听者是将其整合一起，作为一个整体加工，而不是分开加工(Serafine, Davidson, Crowder, & Repp, 1986)。在本研究中，当音乐带有歌词时，听者需要整合旋律与歌词，并可能将其作为一个整体与情绪面孔的一致性进行匹配，由此诱发了 LPC 效应。然而，需要说明的是，尽管听者对带有歌词音乐情绪信息的加工滞后于无歌词音乐(体现为 LPC 效应和 N400 效应潜伏期的差异)，但启动效应的出现意味着听者能正确理解并加工带有歌词与无歌词音乐所传达的情绪信息。

作为两个独立的脑电成分，研究者普遍认为 N400 与 LPC 代表了不同的认知加工过程(Ibáñez et al., 2010; Juottonen et al., 1996; Rohaut & Naccache, 2017)。同时，已有研究表明，听者对带有歌词与无歌词音乐情绪信息的加工可能涉及不同的神经机制(Brattico et al., 2011)。因此，虽然本研究认为 N400 和 LPC 在一定程度上均反映了整合加工，但这两个成分所代表的整合加工应当存在一定差异。基于情绪的起源(origin)，情绪双模型理论(emotion duality model)认为，在个体对刺激做出情绪反应的过程中，存在两个评价机制：自动评价系统(the automatic evaluating system)和反思性评价系统(the reflective evaluating system) (Jarymowicz & Imbir, 2015)。反思性反应的产生需要以言语为基础，没有言语的参与则无法产生这种反应(Imbir, Spustek, & Żygierewicz, 2016)。是否带有语言信息恰恰是本研究中带有歌词与无歌词音乐的差异。因此，本文的 LPC 可能反映的是一种反思性的整合加工，而 N400 可能反映的是一种相对自动化的整合加工。

通过两个实验，本研究发现，歌词的参与导致了听者对音乐情绪加工的滞后，该结果验证了音乐哲学的观点。在音乐哲学家看来，纯音乐(无歌词音乐)在传达情绪方面比带有歌词音乐更迅速更直接(于润洋，2000；张前，王次炤，1992)，这是因为，

语言通过命题系统(propositional system)明确地传达情绪信息(Erickson, 2005; Jankélévitch & Abbate, 2003)，而纯音乐正是没有类似语言的语义，使其在传递情绪过程中，可以省略命题系统中的翻译(translation)过程，由此导致纯音乐能够更迅速更直接地传递情绪信息。

从语言与音乐情绪加工的关系来看，与已有行为研究结果一致(Ali & Peynircioğlu, 2006; Mori & Iwanaga, 2013; Stratton & Zalanowski, 1994)，本研究结果表明歌词对音乐情绪加工具有影响。应该指出的是，本研究是通过比较有/无歌词条件下的音乐情绪加工，探究歌词的影响效应。尽管本研究的实验设计无法直接回答语言与音乐在情绪加工方面是否存在交互的问题，但是，本研究结果暗示了语言与音乐的情绪加工可能共享特定的机制。具体来说，与无歌词条件不同，带有歌词音乐条件诱发了 LPC 效应，该 LPC 效应主要缘于歌词的介入。就带有歌词的条件而言，歌词与旋律是同时呈现，既然歌词影响其音乐情绪加工，则暗示了语言与音乐的情绪加工可能在此存在交互。未来研究需要对此进行进一步验证。

5 结论

本研究表明，无论是否带有歌词，听者都能加工音乐所传达的情绪信息。但是，无歌词音乐条件在 250~450 ms 时间窗口诱发了 N400 效应，而带有歌词音乐条件在 500~700 ms 时间窗口诱发了 LPC 效应，该结果表明歌词影响了大脑加工音乐情绪的时间进程，听者对带有歌词音乐情绪的加工滞后于无歌词音乐。本研究在一定程度上为音乐与语言关系的探究提供了依据。

参 考 文 献

Ali, S. O., & Peynircioğlu, Z. F. (2006). Songs and emotions: Are lyrics and melodies equal partners? *Psychology of Music, 34*(4), 511–534.

Behrens, G. A., & Green, S. B. (1993). The ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychology of Music, 21*(1), 20–33.

Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., & Tervaniemi, M. (2011). A functional MRI study of happy and sad emotions in music with and without lyrics. *Frontiers in Psychology, 2*, 308.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5*(1), 34–44.

Daltrozzo, J., & Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience, 21*(10), 1882–1892.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. London, UK: John Murray.

Erickson, D. A. (2005). *Language, ineffability and paradox in music philosophy* (Unpublished doctorial dissertation). Simon Fraser University.

Franco, F., Chew, M., & Swaine, J. S. (2017). Preschoolers' attribution of affect to music: A comparison between vocal and instrumental performance. *Psychology of Music, 45*(1), 131–149.

Galizio, M., & Hendrick, C. (1972). Effect of musical accompaniment on attitude: The guitar as a prop for persuasion. *Journal of Applied Social Psychology, 2*(4), 350–359.

Goerlich, K. S., Witteman, J., Aleman, A., & Martens, S. (2011). Hearing feelings: Affective categorization of music and speech in alexithymia, an ERP study. *Plos One, 6*(5), e19501.

Goerlich, K. S., Witteman, J., Schiller, N. O., Van -Heuven, V. J., Aleman, A., & Martens, S. (2012). The nature of affective priming in music and speech. *Journal of Cognitive Neuroscience, 24*(8), 1725–1741.

Gong, X., Huang, Y-X., Wang, Y., & Luo, Y-J. (2011). Revision of the Chinese facial affective picture system. *Chinese Mental Health Journal, 25*(1), 40–46.

[龚栩，黄宇霞，王妍，罗跃嘉. (2011). 中国面孔表情图片系统的修订. *中国心理卫生杂志, 25*(1), 40–46.]

Hailstone, J. C., Omar, R., Henley, S. M. D., Frost, C., Kenward, M. G., & Warren, J. D. (2009). It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *Quarterly Journal of Experimental Psychology, 62*(11), 2141–2155.

Herring, D. R., Taylor, J. H., White, K. R., & Crites, S. L. (2011). Electrophysiological responses to evaluative priming: The LPP is sensitive to incongruity. *Emotion, 11*(4), 794–806.

Hu, X., Downie, J. S., & Ehmann, A. F. (2009, October). *Lyric text mining in music mood classification.* Paper presented at the International Society for Music Information Retrieval Conference, Kobe, Japan.

Ibáñez, A., Manes, F., Escobar, J., Trujillo, N., Andreucci, P., & Hurtado, E. (2010). Gesture influences the processing of figurative language in non-native speakers: ERP evidence. *Neuroscience Letters, 471*(1), 48–52.

Imbir, K. K., Spustek, T., & Żygierewicz, J. (2016). Effects of valence and origin of emotions in word processing evidenced by event related potential correlates in a lexical decision task. *Frontiers in Psychology, 7*, 271.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception, 26*(3), 195–204.

Jankélévitch, V., & Abbate, C. (2003). *Music and the ineffable*. Princeton, NJ: Princeton University Press.

Jarymowicz, M. T., & Imbir, K. K. (2015). Toward a human emotions taxonomy (based on their automatic vs. reflective origin). *Emotion Review, 7*(2), 183–188.

Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research, 4*(2), 99–107.

Kamiyama, K. S., Abla, D., Iwanaga, K., & Okanoya, K. (2013). Interaction between musical emotion and facial expression as measured by event–related potentials. *Neuropsychologia, 51*(3), 500–505.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7*(3), 302–307.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463–470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*(1), 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading between the lines: Event-related brain potentials during natural sentence processing. *Brain and Language, 11*(2), 354–373.

Laurier, C., Grivolla, J., & Herrera, P. (2008, May). *Multimodal music mood classification using audio and lyrics*. Paper presented at the Seventh International Conference on Machine Learning and Applications, Copenhagen, Denmark.

Levman, B. G. (1992). The genesis of music and language. *Ethnomusicology, 36*(2), 147–170.

Mithen, S. J. (2006). The singing Neanderthals: The origins of music, language, mind and body. *Cambridge Archaeological Journal, 59*(2), 54.

Mori, K., & Iwanaga, M. (2013). Pleasure generated by sadness: Effect of sad lyrics on the emotions induced by happy music. *Psychology of Music, 42*(5), 643–652.

Morton, J. B., J., & Trehub, S. E. (2007). Children's judgements of emotion in song. *Psychology of Music, 35*(4), 629–639.

Perlovsky, L. (2010). Musical emotions: Functions, origins, evolution. *Physics of Life Reviews, 7*(1), 2–27.

Rohaut, B., & Naccache, L. (2017). Disentangling conscious from unconscious cognitive processing with event–related EEG potentials. *Revue Neurologique, 173*(7–8), 521–528.

Serafine, M. L., Davidson, J., Crowder, R. G., & Repp, B. H. (1986). On the nature of melody-text integration in memory for songs. *Journal of Memory and Language*, *25*(2), 123–135.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research, 14*(2), 228–233.

Sousou, S. D. (1997). Effects of melody and lyrics on mood and memory. *Perceptual & Motor Skills, 85*(1), 31–40.

Steinbeis, N., & Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *Journal of Cognitive Neuroscience, 23*(3), 604–621.

Stratton, V. N., & Zalanowski, A. H. (1994). Affective impact of music vs. lyrics. *Empirical Studies of the Arts, 12*(2), 173–184.

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences, 109*(46), 19027–19032.

Wang, Y., & Qin, Z. (2016). Affective priming by simple geometric shapes: Evidence from event-related brain potentials. *Frontiers in Psychology, 7*, 917.

Werheid, K., Alpay, G., Jentzsch, I., & Sommer, W. (2005). Priming emotional facial expressions as evidenced by event–related brain potentials. *International Journal of Psychophysiology, 55*(2), 209–219.

Yu, R. Y. (2000). *An introduction to modern western music philosophy*. Changsha, China: Hunan Education Press.

[于润洋. (2000). *现代西方音乐哲学导论*. 长沙: 湖南教育出版社.]

Zhang, Q., Kong, L., & Jiang, Y. (2012). The interaction of arousal and valence in affective priming: Behavioral and electrophysiological evidence. *Brain Research*, *1474*, 60–72.

Zhang, Q., Li, X. H., Gold, B. T., & Jiang, Y. (2010). Neural correlates of cross-domain affective priming. *Brain Research, 1329*, 142–151.

Zhang, Q., & Wang, C. Z. (1992). *Foundation in musical aesthetics*. Beijing, China: People's Education Press.

[张前, 王次炤. (1992). *音乐美学基础*. 北京: 人民音乐出版社.]

Zhang, Q., Lawson, A., Guo, C. Y., & Jiang, Y. (2006). Electrophysiological correlates of visual affective priming. *Brain Research Bulletin, 71*(1-3), 316–323.

Effects of lyrics on the processing of musical emotion: Behavioural and ERP study

ZHANG Weixia[1]; WANG Wanqi[2]; ZHOU Linshu[3]; JIANG Cunmei[3]

(1 *Department of Psychology, College of Education, Shanghai Normal University, Shanghai* 200234, *China*)

([2] *School of Finance and Business, Shanghai Normal University, Shanghai* 200234, *China)*

([3] *Music College, Shanghai Normal University, Shanghai* 200234, *China*)

Abstract

Music and language are unique to the human beings. It has been suggested that music and language have a common origin as an emotional protolanguage. The development of socialisation has resulted in the development of language into a symbolic communication system with explicit semantics. By contrast, music has become an important means of emotional expression. However, whether language with explicit semantics affects the emotional processing of music remains uncertain. Given that songs contain melody and lyrics, previous behavioural studies have focused on songs to analyse the influence of lyrics on the processing of musical emotion. However, several studies have also shown the influence of lyrics, although such findings are relatively

contradictory.

Thus, the current study used behavioural and electrophysiological measurements to investigate the impact of lyrics on the processing of musical emotion. Experiment 1 analysed whether the emotional connotations in music with and without lyrics could be perceived by listeners at the behavioural level. Experiment 2 further investigated whether there are different neural responses to emotions conveyed by melodies with and without lyrics.

A cross-modal affective priming paradigm was used in Experiments 1 and 2, in which musical excerpts served as the prime and emotional faces as target. To avoidthe impact of familiarity, 120 musical stimuli were selected from European opera. Each was sung by a vocalist with and without lyrics, thereby resulting in 240 musical stimuli in two versions as potential prime stimuli. A total of 160 facial expressions affectively congruent or incongruent with the preceding musical stimuli were selected as potential target stimuli. Three pre-tests were conducted to ensure the validity of the stimuli. Eventually, 60 musical stimuli for each music version were selected as the prime stimuli, whilst 120 images were used as the target stimuli, thereby resulting in 240 music–image pairs. To ensure that each stimulus appears only once for each participant, two lists were prepared using a Latin square design. Each prime and target was presented in either the congruent or incongruent condition within each list. Thus, each list comprised 120 trials, with 30 trials in each condition. During the experiment, the two lists were equally distributed across the participants. A total of 40 healthy adults participated in Experiment 1. They were asked to judge as quickly and accurately as possible whether the emotion of the target was happy or sad. The accuracy and reaction time were collected. Meanwhile, 20 healthy adults participated in Experiment 2. They were required to judge whether the emotion between music and image was congruent or incongruent whilst their EEG waveforms were recorded. ERPs were analysed and compared between conditions at the time windows of 250~450 ms and 500~700 ms after the onset of the target.

The Experiment 1 results showed that when faces were primed by music either with or without lyrics, the participants responded faster and more accurately under affectively congruent condition compared with affectively incongruent condition. This finding indicated that the emotional connotations in music with and without lyrics could both be perceived. The ERP results in Experiment 2 showed that distinct neural mechanisms were activated by music with and without lyrics. Specifically, when faces were primed by music without lyrics, a larger N400 was elicited in response to affectively incongruent pairs than to affectively congruent pairs at the time window of 250~450 ms. However, when faces were primed by music with lyrics, a more positive LPC was observed in response to the affectively incongruent pairs than to the affectively congruent pairs at 500~700 ms. This finding confirms the results of Experiment 1, thereby suggesting that the emotion conveyed by music with and without lyrics could be perceived by the listeners. Moreover, the emotional processing between music with and without lyrics differs in the time course of neural processing. That is, the emotional processing of music with lyrics lagged behind that of music without lyrics.

In conclusion, the present results suggest that the neural processing of emotional connotations in music without lyrics preceded that of music with lyrics, although the emotional connotations conveyed by music with and without lyrics could both be perceived. These findings also supported theory of musical philosophy, which suggests that music without lyrics can express emotion more immediately and more directly than music with lyrics owing to the lack of “translation” from the propositional system. On the other hand, considering that lyrics influenced the time course of emotional processing in music with lyrics, our results also provide evidence that the emotional processing of music and language may share neural resources to some extent.

Key words　musical emotion; language; lyrics; N400; LPC